The International Conference on Management, Business, Economics, Law and Technology (COMBELT-2025)

August 01st, 2025, Danang, Vietnam

The Imperative of Retrieval Augmented Generation (RAG) for ESG Compliance and Financial Regulation: A Qualitative Thematic Analysis of the Research Landscape

N. Long Ha[a✉], S. Huong Do[a], Quan Truong Tan[a] Van Phuc Le[b]

[a]*University of Economics, Hue University, 99 Ho Dac Di, Hue City 54000, Vietnam*

[b]*Foreign Trade University, 90 Chua Lang, Hanoi 100000, Vietnam*

ABSTRACT

The rising prominence of Environmental, Social, and Governance (ESG) criteria, driven by regulatory mandates and stakeholder expectations, is reshaping corporate strategies and financial markets. This shift creates voluminous, complex, and dynamic ESG data and disclosures. Traditional methods and standalone Large Language Models (LLMs) struggle with the specificity and evolving nature of ESG regulatory texts. Employing a systematic literature review (Web of Science, Scopus) and qualitative thematic analysis, this paper synthesizes research to articulate the critical need for Retrieval Augmented Generation (RAG) in the ESG Regulatory Technology (RegTech) domain. Our analysis delineates LLM shortcomings in ESG compliance and supervision, such as factual inaccuracies, static knowledge, and limited contextual understanding of complex regulatory documents. We posit that RAG, by grounding LLM outputs in verifiable, current external knowledge like regulatory texts and corporate disclosures, offers a transformative advancement. This approach is critical for robust ESG RegTech solutions, enabling effective compliance monitoring and informed, evidence-based decision-making in a sustainability-focused economy.

Keywords: ESG; Financial Regulation; Large Language Models (LLMs); RegTech; Retrieval Augmented Generation (RAG); Qualitative Analysis.

1. Introduction

✉ Corresponding author. Tel.: +84935626575.
Email address: hnlong@hueuni.edu.vn

The global financial and corporate landscape is transforming as Environmental, Social, and Governance (ESG) criteria become integral to investment, strategy, and regulatory oversight, reflecting their link to long-term value, risk mitigation, and societal impact. This has exponentially increased complex ESG data, disclosures, and mandates under frameworks like TCFD, GRI, SFDR, and CSRD. This proliferation of often unstructured, diverse information poses considerable challenges for effective analysis and verifiable compliance (Abdulrahman Bagais et al., 2024). To address these ESG complexities, Regulatory Technology (RegTech) has emerged, particularly for financial institutions. RegTech employs advanced technology to help organizations navigate dynamic ESG regulations, streamline compliance, manage risks, and improve reporting transparency by automating traditionally manual processes.

Concurrently, advancements in artificial intelligence (AI), particularly sophisticated Large Language Models (LLMs), show transformative potential in finance and regulatory compliance (Braun and Clarke, 2006). LLMs excel at processing vast text datasets. However, standalone LLMs have significant limitations in specialized domains: their knowledge is static, reflecting only their last training data, and they can "hallucinate"—generating plausible but incorrect information (Du et al., 2025). These issues are particularly acute for ESG regulatory compliance, where accuracy, currency, and traceability are paramount.

Retrieval Augmented Generation (RAG) offers a promising architecture to mitigate these LLM weaknesses by grounding outputs in external, verifiable knowledge sources consulted before generation (Fan et al., 2024). This is vital for ESG regulatory compliance, which demands factual precision and current information. The synergy of ESG's data demands, RegTech's objectives, and RAG-LLMs' capabilities represents a significant advancement (Huang et al., 2023). While LLMs can process unstructured ESG text, their static knowledge and hallucination risks remain problematic. RAG addresses these issues by connecting LLMs to curated, current external knowledge bases, creating a more robust, multi-layered approach (Huang et al., 2023).

The dynamic ESG regulatory environment demands adaptable technological solutions. RAG's ability to update its knowledge base by refreshing external data sources—without costly LLM retraining—makes it a more agile and future-proof solution for ESG RegTech than static LLMs or fine-tuned models requiring intensive updates (Kim et al., 2024). This paper, through a systematic literature review (SLR) of Web of Science and Scopus databases and a subsequent qualitative thematic analysis, explores the imperative for RAG in ESG compliance and financial regulation. We aim to delineate LLM limitations in ESG regulatory adherence and demonstrate, through synthesized scholarly discourse, how RAG can provide more reliable, transparent, and effective solutions.

This study is guided by the following central research questions:

RQ1: What are the current applications and documented limitations of Large Language Models (LLMs) in the context of Environmental, Social, and Governance (ESG)-related financial regulation, supervision, and compliance (RegTech), as identified in the peer-reviewed literature from Web of Science (WoS) and Scopus?

RQ2: Which specific challenges (e.g., interpreting nuanced regulatory language, ensuring up-to-date knowledge of evolving regulations, verifying compliance from voluminous and diverse corporate reports, mitigating hallucination risks) are identified as critical in ESG RegTech that standalone LLMs struggle to address effectively, according to the selected scholarly literature?

RQ3: How can Retrieval Augmented Generation (RAG) theoretically and practically enhance the capabilities of LLMs for ESG regulatory analysis, compliance checking, and reporting verification, based on insights, conceptualizations, and early evidence presented in the reviewed academic literature?

RQ4: What is the current state of peer-reviewed research (as indexed in WoS and Scopus) regarding the explicit application, empirical evaluation, or robust conceptualization of RAG for ESG RegTech, compliance, and financial regulation, and what are the key underexplored areas or significant research gaps that warrant future scholarly investigation?

This study critically examines academic literature to establish the imperative for RAG in AI-driven ESG compliance and financial regulation. Its objectives are to systematically review and synthesize peer-reviewed literature (WoS, Scopus) on LLMs and RAG in this domain; conduct a qualitative thematic analysis to assess the "need" for RAG by exploring LLM shortcomings; delineate RAG's advantages regarding accuracy, currency, transparency, and contextual understanding; map current RAG research in ESG regulation, identifying explorations and gaps; and propose RAG as a pivotal technology for enhancing AI-driven ESG regulatory processes. The paper is organized as follows: Section 2 reviews literature on AI and LLMs in ESG regulatory dimensions. Section 3 details the SLR and qualitative thematic analysis methodology. Section 4 will present findings. Section 5 discusses these findings' significance and implications. Section 6 concludes, summarizing arguments and proposing future research.

2. Literature review

The intersection of Artificial Intelligence (AI), particularly Large Language Models (LLMs), and Environmental, Social, and Governance (ESG) considerations within the financial regulatory landscape is a dynamic and critical research area. This section reviews existing literature to map the state-of-the-art, focusing on AI and LLM roles in ESG-related financial regulation and compliance, thereby establishing the context for RAG's emerging imperative.

2.1. The expanding role of AI in ESG financial regulation and compliance

Artificial Intelligence (AI) is increasingly utilized in finance for complex decisions, large-scale data analysis, pattern identification, and predictive insights. In ESG, AI enhances sustainability assessment, risk management, and compliance, improving regulatory efficiency, accuracy, and transparency (Lee et al., 2025). Regulatory Technology (RegTech) leverages AI for monitoring and Supervisory Technology (SupTech), with authorities using AI for risk management, automated reporting, market surveillance, and fraud detection (Li et al., 2025). For ESG, AI automates data tracking/analysis, aiding regulatory enforcement, corporate governance, and developing dynamic risk-anticipation indices. AI-augmented decision-making is thus vital for market monitoring and risk management, as SupTech increasingly addresses ESG risk assessments (Lim, 2024).

Natural Language Processing (NLP), a key AI subfield in finance, processes textual data, extracting information from extensive, unstructured ESG reports and financial texts, often via deep learning (Lim, 2024). Core NLP methods (e.g., tokenization, named entity recognition/NER) structure ESG narratives; its use in sustainability report analysis is active research. Advanced NLP models (e.g., BERT, FinBERT, ClimateBERT) are dominant in ESG/AI finance for extracting nuanced insights from ESG texts (Alahira et al., 2024; Olaiya et al., 2024; Ogunyemi, 2023). Combined with RAG and agentic workflows, NLP also powers chatbots for navigating complex regulatory information. Ultimately, NLP automates report analysis, extracts actionable insights, and supports informed ESG decision-making.

2.2. Large language models in ESG regulation and compliance

Large Language Models (LLMs), such as the GPT series, Llama, and Mistral, have demonstrated remarkable capabilities in processing and analyzing text at scale, positioning them as powerful tools for ESG reporting, regulatory compliance, and data-driven decision-making in finance (Parikh and Penfield, 2024).

2.2.1. Potential of LLMs for ESG RegTech and compliance

LLMs are emerging as potent instruments for enhancing operational risk management and streamlining compliance within the financial sector due to their sophisticated analytical capabilities (Parikh and Penfield, 2024). In ESG contexts, LLMs offer significant potential to simplify reporting by automating the extraction of structured information (e.g., metrics, policy commitments) from unstructured sources like sustainability reports and regulatory documents (Abdulrahman Bagais et al., 2024).

Generative AI, including LLMs, can assist companies in addressing complex ESG requirements, such as “double materiality” assessments under CSRD, by deriving insights from NLP analysis of diverse data (Braun and Clarke, 2006). LLMs can also support strategic alignment with evolving ESG regulatory frameworks and facilitate the structuring of unstructured data (e.g., emails, meeting notes) for improved analysis in financial supervision.

2.2.2. Critical limitations of LLMs in the demanding ESG regulatory domain

Despite their potential, standalone LLMs exhibit significant limitations in specialized, high-stakes domains like ESG regulation. A primary weakness is their tendency to “hallucinate”—generating plausible but factually incorrect outputs (Du et al., 2025). This is particularly problematic where accuracy is non-negotiable. The “black-box” nature of LLMs and hallucination risks are major concerns for ESG applications.

Standard LLMs often require domain adaptation or fine-tuning for optimal performance on specific ESG tasks, as general-purpose models may not capture necessary nuances (Page et al., 2021). The static knowledge of LLMs (based on their last training date) prevents access to the latest regulatory changes without retraining. Furthermore, the length of many ESG reports and regulatory documents can exceed LLM context window limitations, hindering comprehensive analysis and incurring substantial API costs for larger models (Sun et al., 2024). The opacity of complex AI models also poses risks associated with inaccurate predictions and lack of explainability (Braun and Clarke, 2006).

2.3. Retrieval Augmented Generation

Retrieval Augmented Generation (RAG) has emerged as an architectural approach to enhance LLM capabilities by integrating external, verifiable knowledge, directly addressing key LLM limitations such as hallucination and static knowledge.

2.3.1. RAG Principles and Potential in ESG RegTech

RAG augments LLMs by enabling them to consult an authoritative external knowledge base before generating a response, a crucial distinction from standalone LLMs that rely solely on their internal, parametric memory. This integration of external, non-parametric memory allows RAG systems to produce more reliable and contextually relevant outputs. The process typically involves two main phases: first, a retrieval mechanism searches an external knowledge base (e.g., ESG regulations, corporate reports) for information segments relevant to a user query, often using vector embeddings for semantic similarity.

Second, this retrieved information is combined with the original query to form an augmented prompt, which the LLM then uses to generate a comprehensive, factually grounded response. The creation and maintenance of a high-quality external knowledge base are vital for RAG's effectiveness. By dynamically retrieving information, RAG addresses LLM limitations like hallucinations and outdated knowledge, thereby improving accuracy without extensive LLM fine-tuning for new information (Abdulrahman Bagais et al., 2024).

RAG systems are increasingly used in knowledge-intensive domains like law and finance to reduce LLM hallucinations and provide factual knowledge beyond their training data. This paradigm, often combined with techniques like In-Context Learning (ICL), leverages LLMs' semantic understanding for more effective structured information extraction. For instance, RAG has improved GPT-4's climate-related question answering by augmenting queries with IPCC report data (Thimm and Rasmussen, 2024). Well-designed RAG pipelines demonstrably reduce hallucinations and enhance task performance, making RAG promising where factual accuracy and currency are essential.

2.3.2. Conceptual and Early Applications of RAG in ESG Compliance and Supervision

RAG pipelines are being deployed in sustainability research to add domain-specific context to LLM prompts, enhancing tasks like information extraction from corporate reports and checking alignment with disclosure rules (Abdulrahman Bagais et al., 2024). Researchers are exploring RAG for extracting sustainability data using models like Llama for fine-grained assessments and for benchmarking language model performance in processing sustainability reports (Vaghefi et al., 2023).

In the legal ESG domain, RAG is being investigated for tasks like determining the applicability of ESG regulations to companies and ESG regulation question-answering (Sun et al., 2024). Modern RAG pipelines are developed for these tasks to reduce hallucinations and improve performance, with early experiments showing potential to support practitioners in ESG compliance. The aim is to create IT tools using RAG to assist experts, increasing efficiency in understanding and complying with ESG regulations, though such systems are intended to support, not replace, human expertise. RAG architectures have also been utilized for broader ESG data extraction and analysis from corporate reports (Abdulrahman Bagais et al., 2024). Furthermore, NLP combined with RAG and agentic workflows is conceptualized for regulatory authorities to build internal chatbots or public information retrieval systems (Parikh and Penfield, 2024).

3. Methodology

This study investigates the need for Retrieval Augmented Generation (RAG) in ESG compliance and financial regulation. Our methodology integrates a systematic literature review (SLR) of Web of Science (WoS) and Scopus with a qualitative thematic analysis of selected studies. This section details this dual-phase protocol for addressing research questions (RQ1-RQ4).

3.1. Research Design: A Two-Phase Approach

Our research is structured in two distinct yet interconnected phases, ensuring both breadth in coverage and depth in analysis:

Phase 1: Systematic Literature Review (SLR): The primary objective of this initial phase is the systematic and comprehensive identification, meticulous screening, and rigorous selection of relevant peer-reviewed academic literature. The SLR is specifically focused on articles and conference proceedings indexed in the Web of Science (WoS) and Scopus databases. This phase aims to comprehensively map the current state of published research concerning the application, potential, and limitations of LLMs, and the emerging role or necessity of RAG, within the specific domain of ESG financial regulation, supervision, and compliance. The output of this phase is a carefully curated and well-defined corpus of scholarly literature. This corpus provides the foundational dataset upon which the subsequent qualitative thematic analysis is built and is instrumental in gathering the primary evidence required to address RQ1, RQ2, RQ3, and RQ4.

Phase 2: Qualitative Thematic Analysis: Building directly upon the literature corpus identified and selected in Phase 1, this second phase involves an in-depth qualitative thematic analysis of the full textual content of the chosen studies. The goal of this phase is to move beyond a merely descriptive summary or aggregation of the literature. Instead, it focuses on an interpretive synthesis of the evidence, arguments, conceptual discussions, and empirical findings presented within the selected papers. This qualitative analysis is specifically designed to identify, analyze, and report patterns (themes) that articulate the nuanced limitations of standalone LLMs and establish the specific, evidence-based needs that RAG can fulfill in the complex ESG regulatory and compliance contexts. This interpretive approach provides the necessary nuanced and rich insights required for a thorough and insightful exploration of RQ1, RQ2, RQ3, and RQ4.

3.2. Phase 1: Systematic Literature Review Protocol

The SLR was conducted with adherence to established academic best practices and recognized guidelines for systematic reviews to ensure methodological rigor, transparency, and the potential for replicability.

3.2.1. Adherence to Standards:

The review process was guided by PRISMA principles (Villacampa-Porta et al., 2025), adapted for a review focusing on technological applications, conceptual needs, and qualitative synthesis.

3.2.2. Search strategy:

The literature search exclusively used the Web of Science (WoS) Core Collection and Scopus. These major multidisciplinary databases were chosen for their comprehensive coverage and robust search functionalities, facilitating targeted and reproducible queries. The search targeted English-language articles across the full range of available publications to ensure broad coverage of relevant studies.

Search String Construction: A precise search string was developed to capture literature on LLMs/RAG and ESG within financial regulation, RegTech, supervision, and compliance: *("Retrieval Augmented Generation" OR RAG OR LLM OR "Large Language Model") AND (ESG AND (RegTech OR "Financial Regulation" OR "Financial Supervision" OR Compliance))*

3.2.3. Study Selection Criteria:

Peer-reviewed WoS/Scopus literature underwent multi-stage screening (title/abstract, then full-text). Inclusion required substantive discussion of LLMs/RAG related to ESG compliance, financial regulation, RegTech, or supervision. Excluded were non-indexed/non-peer-reviewed items, general AI papers lacking specific ESG regulatory focus, peripheral ESG mentions, non-English articles, and duplicates.

3.2.4. Data Extraction for Systematic Literature Review

For each academic study that met all inclusion criteria and passed the full-text review, a structured data extraction form (or spreadsheet) was developed and utilized. This form was designed to systematically capture key information pertinent to the research questions. Extracted data points included: full bibliographic details (authors, title, year, journal/conference, DOI), specific ESG regulatory area or focus (e.g., climate disclosure, anti-greenwashing, SFDR, CSRD), LLM or RAG techniques discussed or applied, stated benefits or potentials of the AI approach, identified limitations of LLMs in the ESG regulatory context, explicit arguments for or applications of RAG (if present), main outcomes or conclusions of the study, and noted future research directions relevant to the RQs of this paper.

3.3. Phase 2: Qualitative thematic analysis of selected literature

The final corpus of full-text scholarly articles and conference papers, meticulously selected through the SLR process as described above, constituted the primary dataset for the subsequent in-depth qualitative thematic analysis. This phase aimed to synthesize the identified literature to build a rich, interpretive understanding of the scholarly discourse surrounding the imperative for RAG in ESG RegTech.

3.3.1. Analytical focus

The core analytical objective of this qualitative phase was to move beyond a mere aggregation or descriptive summary of the findings from individual papers. Instead, the focus was on identifying, meticulously analyzing, and systematically reporting cross-cutting patterns, salient arguments, and overarching themes that emerge from the collective scholarly discourse within the selected literature. These themes were specifically sought to articulate with clarity and depth: (a) the nuanced limitations, specific challenges, and practical shortcomings of standalone LLMs when applied to the multifaceted tasks of ESG regulatory analysis and compliance verification (RQ1, RQ2), and (b) the specific, unmet needs within ESG RegTech, the conceptual arguments presented, and the potential practical advantages that position RAG as a necessary, beneficial, and potentially transformative advancement in this domain (RQ3).

3.3.2. Data for qualitative analysis

The primary data comprised the complete textual content—including introductions, literature reviews, methodologies, findings, discussions, and conclusions—of all peer-reviewed journal articles and full conference papers that were selected and deemed eligible through the SLR screening process described in Section 3.2.

3.3.3. Thematic analysis process

An inductive thematic analysis approach was adopted, drawing methodological guidance from the widely recognized and well-established framework proposed by Braun & Clarke (2006) for conducting thematic analysis in qualitative research (Webersinke et al., 2021; Moodaley and Telukdarie, 2023). This iterative and reflexive process involved the following key steps:

Familiarization with the data: This foundational step involved intensive and repeated close reading of all selected articles to achieve a deep, holistic, and nuanced understanding of their content, core arguments, methodologies employed, reported findings, and broader implications related to the use of LLMs, the potential or necessity of RAG, and the specific challenges within ESG regulation and compliance.

Generating initial codes: This step involved systematically identifying and labeling relevant segments of text (ranging from phrases and sentences to entire paragraphs) across the entire dataset. Codes were developed to capture explicit discussions of LLM capabilities, their limitations in ESG regulatory contexts, specific compliance challenges encountered, the described or potential role and architecture of RAG systems, arguments supporting its necessity or benefits, and any reported evidence of its application or effectiveness.

Searching for themes: This interpretive step involved collating the generated codes into potential overarching themes. It required looking for recurring patterns of meaning, common or contrasting arguments, areas of scholarly consensus or ongoing debate, and significant insights that emerged across the coded data, particularly those directly addressing the research questions concerning the "need for RAG."

Reviewing themes: This critical step involved evaluating the coherence, distinctiveness, and comprehensiveness of the initially identified potential themes. Themes were reviewed against both the coded extracts and the entire dataset to ensure they accurately and meaningfully represented the data. Themes were refined, merged if overly similar, subdivided if too broad, or discarded if not sufficiently supported by the data.

Defining and naming themes: This step focused on clearly articulating the scope, central organizing concept, and essential essence of each finalized theme. It involved developing concise, informative, and analytically robust names for each theme and writing detailed narrative descriptions that encapsulated what each theme represented in the context of the research questions.

Producing the report: The final step involved weaving the analyzed and defined themes into a coherent, compelling, and well-structured analytical narrative. This narrative, which forms the core of Section 4.2 of this paper, is supported by illustrative quotations, synthesized evidence, and interpretive insights drawn directly from the selected body of scholarly literature.

3.4. Quality assurance and rigor

Quality assurance was integral to both research phases. For the SLR, methodological transparency was ensured through explicit documentation of the search strategy, selection criteria, and transparent reporting of results (e.g., via a PRISMA diagram). The credibility of the qualitative thematic analysis was enhanced by strategies such as maintaining a detailed audit trail of coding and theme development, ensuring inter-coder reliability if applicable, and grounding all themes firmly in textual evidence from the selected literature.

4. Findings

4.1. Overview of literature from WoS and Scopus

The systematic literature review commenced with a comprehensive search of the Web of Science (WoS) Core Collection and Scopus databases, adhering to the search strategy detailed in Section 3.2.2. The initial search query yielded a total of 65 records, with 64 unique records identified from Scopus and 1 record from Web of Science. One record retrieved from Web of Science was found to be a duplicate of a record already present in the Scopus results and was subsequently removed. This left a total of 64 unique records for the screening process.

The screening process was conducted in multiple stages:

1. **Title Screening:** The 64 unique records underwent an initial title screening. Based on the predefined inclusion and exclusion criteria (Section 3.2.3), 33 records were excluded at this stage. This resulted in 31 records proceeding to abstract screening.
2. **Abstract Screening:** The abstracts of these 31 records were then meticulously reviewed. This led to the exclusion of a further 19 records that did not meet the study's scope. Consequently, 12 records were deemed potentially eligible and were selected for full-text assessment.
3. **Full-Text Assessment:** The full texts of these 12 articles were retrieved and thoroughly assessed for eligibility. During this stage, 3 articles were excluded as they were found to be out of scope, primarily because they did not substantively address the intersection of RAG and ESG within the financial regulatory or compliance context.

Therefore, a final corpus of 9 studies met all inclusion criteria and were included for the in-depth qualitative thematic analysis. The flow of studies through these identification and screening phases is typically illustrated with a PRISMA flow diagram (cf. Fig. 1), which visually represents the number of records at each stage of the review process.

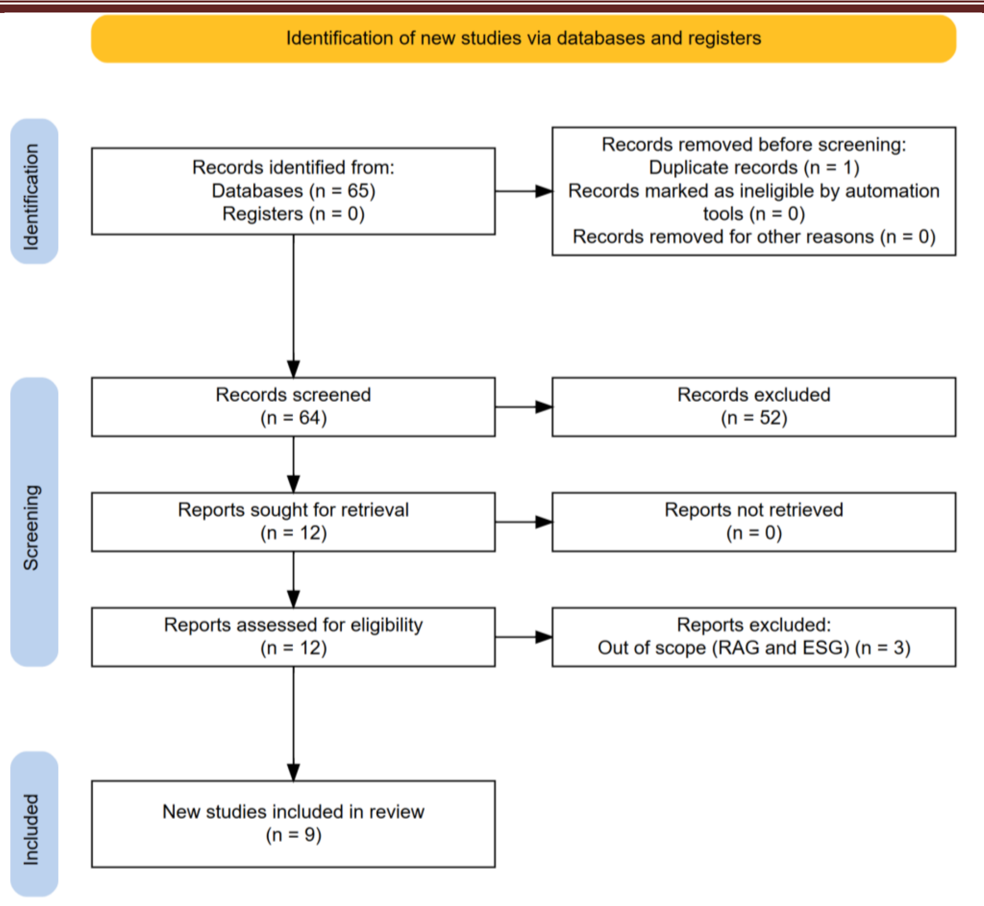

Fig. 1. PRISMA flow diagram of literature screening and selection

The analysis of the nine included studies reveals a highly contemporary research landscape, with all publications from 2023-2025, predominantly 2024-2025, underscoring emergent scholarly interest in AI for ESG regulation. This research exhibits geographical diversity in empirical applications (Europe, North America, Asia) alongside global reviews, spanning interdisciplinary fields like environmental science, information systems, and AI. Key themes include enhancing ESG disclosure, climate risk reporting, automated information extraction, greenwashing detection, and AI's role in compliance and risk management. All studies utilize Large Language Models (LLMs) or advanced NLP for ESG tasks. While explicit Retrieval Augmented Generation (RAG) implementation isn't central in most empirical works, the identified challenges and proposed solutions frequently align with RAG's principles. Notably, Parikh et al. (Sun et al., 2024) describe a system mirroring RAG's functionality. By detailing standalone LLM limitations in handling dynamic, voluminous, and nuanced ESG information, this corpus collectively builds a strong case for the necessity of more robust, verifiable, and context-aware AI systems like RAG, thus underscoring its emerging imperative. A detailed overview of these nine selected studies, summarizing their publication details, geographical focus, key ESG areas addressed, and primary AI/LLM methodologies, is presented in Table 1.

Table 1. Characteristics of Included Peer-Reviewed Articles

Paper ID	Year	Geographical Focus/Context	Publication Venue	Key ESG Area/Challenge Addressed	Primary AI/LLM Focus & RAG Relevance

(Page et al., 2021)	2025	Spanish companies (EU non-financial reporting regulation)	Environmental Sciences Europe	Environmental disclosure quality; Climate risk reporting; Impact of mandatory reporting	NLP, ClimateBERT (LLM), fine-tuned transformer for analyzing disclosures. Sets stage for robust LLM applications in detailed ESG text analysis.
(Yang et al., 2024)	2025	S&P 500 ESG index companies (US context)	Journal of Cleaner Production	Extracting ESG signals from earnings calls; Social impact; Corporate performance and responsibility	NLP, BERT-based summarization, topic modeling for ESG content extraction. Demonstrates LLM use for information extraction from unstructured corporate communications.
(Parikh and Penfield, 2024)	2025	General (Survey)	Information Fusion	Survey of NLP in finance, including ESG & sustainable finance, regulatory compliance monitoring	Comprehensive survey of NLP applications (including LLMs) in finance. Establishes LLM role in ESG finance; RAG could be relevant for advanced QA or compliance aspects discussed.
(Zou et al., 2025)	2025	General (Systematic Literature Review)	Journal of Accounting Literature	LLMs' impact on accounting: ESG disclosure, financial reporting, risk management	SLR on LLMs in accounting. Highlights LLM potential and challenges (data quality, domain adaptation). Mentions "intelligent knowledge bases," hinting at the need for grounded systems.
(Du et al., 2025)	2024	Large production companies (Global/European context via NACE)	Journal of Cleaner Production	Discovery of Green Image Damaging (GID) info; Greenwashing; Environmental non-compliance	ChatGPT (LLM) for retrieving specific GID information. Shows LLM utility in targeted information retrieval, a component of RAG-like processes.
(Braun and Clarke, 2006)	2024	General (Review)	Artificial Intelligence Review	Review of ESG & AI in finance: ESG disclosure, measurement, governance, risk management, responsible AI	Maps AI techniques in ESG finance research. Provides context for advanced AI like RAG, particularly under "Responsible Use of AI" or handling complex "Data."
(Sun et al., 2024)	2024	General (System for corporations)	International Journal of Data Warehousing and Mining	Automatic Question Answering from large ESG reports; Compliance audits; Scope 3 GHG estimation support	Transformer model, LLM for extractive QA from lengthy ESG reports; system locates answer source. Directly addresses a core RAG use case (QA from large documents with source attribution).
(Abdulrahman Bagais et al., 2024)	2024	Real estate industry in Singapore	IEEE Int. Conf. on Data Mining Workshops (ICDMW)	Information extraction from ESG reports; Standardization of ESG metrics; Corporate sustainability reporting	NLP-based automatic extraction algorithm for structuring ESG data. Focuses on information extraction, a process often foundational to or complementary with RAG systems.
(Webersinke et al., 2021)	2023	General (Systematic Literature Review)	Sustainability (Switzerland)	Greenwashing; Sustainability reporting; Role of AI	SLR on AI/ML for greenwashing & sustainability reporting; notes the field as underexplored. Explores AI use, setting the stage for advanced systems to address sophisticated ESG challenges.

4.2. Qualitative thematic analysis

This qualitative thematic analysis synthesizes findings from the nine reviewed WoS and Scopus papers to articulate the imperative for Retrieval Augmented Generation (RAG) in ESG Regulatory Technology (RegTech). This necessity stems from escalating ESG data challenges, critical limitations of standalone Large Language Models (LLMs) in high-stakes regulatory contexts, stringent ESG compliance demands, and RAG's unique ability to bridge these gaps. The analysis is structured to address RQ1, RQ2, and RQ3.

4.2.1. Current LLM Applications and Documented Limitations in ESG Financial Regulation, Supervision, and Compliance (Addressing RQ1)

The reviewed literature shows that Large Language Models (LLMs) are increasingly applied in ESG, financial regulation, and compliance. Applications include ESG information extraction and analysis from unstructured reports and financial texts (Abdulrahman Bagais et al., 2024; Sun et al., 2024; Webersinke et al., 2021; Yang et al., 2024) such as identifying "green image damaging information" (Du et al., 2025). Advanced NLP models like BERT, FinBERT, and ClimateBERT are noted for extracting nuanced insights (Braun and Clarke, 2006; Page et al., 2021). LLMs also demonstrate potential in automating financial and ESG reporting and automatic question answering from large ESG reports (Sun et al., 2024; Zou et al., 2025), and underpin broader NLP applications in finance (Parikh and Penfield, 2024).

Despite these applications, significant limitations persist. A primary concern is their tendency towards factual inaccuracies and "hallucinations"—generating plausible but incorrect information—which is critical in high-stakes regulatory compliance (Du et al., 2025; Parikh and Penfield, 2024; Zou et al., 2025). Their static and outdated knowledge is another major drawback, as LLMs cannot inherently keep pace with the rapidly evolving ESG regulatory landscape (Du et al., 2025). The transparency and verifiability of LLM outputs are also challenged by their "black-box" nature and lack of direct source citation, hindering auditable contexts (Du et al., 2025; Zou et al., 2025). Additionally, many ESG reports exceed standard LLM context windows, necessitating costly models or complex preprocessing (Sun et al., 2024). Finally, general-purpose LLMs often struggle with the complex and nuanced language and specialized terminology of ESG regulations, leading to potential misinterpretations and a lack of domain specificity (Abdulrahman Bagais et al., 2024; Page et al., 2021; Webersinke et al., 2021; Zou et al., 2025).

4.2.2. Critical Challenges in ESG RegTech Ineffectively Addressed by Standalone LLMs (Addressing RQ2)

Standalone Large Language Models (LLMs) face critical challenges in ESG RegTech. They often struggle with interpreting nuanced regulatory language, lacking the specialization to grasp complex details (Abdulrahman Bagais et al., 2024; Zou et al., 2025). A significant hurdle is ensuring up-to-date knowledge of evolving regulations, as LLMs' static knowledge cannot keep pace with the dynamic ESG landscape (Du et al., 2025). Verifying compliance from voluminous and diverse corporate reports is problematic; while LLMs process text, their limitations in handling large documents (Sun et al., 2024), ensuring factual accuracy, and providing verifiable sources impede reliable compliance verification (Abdulrahman Bagais et al., 2024; Du et al., 2025). Mitigating hallucination risks—generating plausible but incorrect information—is crucial due to the high-stakes nature of ESG RegTech outputs (Du et al., 2025). Lastly, effectively handling data heterogeneity and consistency, stemming from diverse ESG reporting frameworks, remains a challenge for standalone LLMs without robust grounding (Abdulrahman Bagais et al., 2024).

4.2.3. RAG's Theoretical and Practical Enhancement of LLMs for ESG Regulatory Analysis and Compliance (Addressing RQ3)

Retrieval Augmented Generation (RAG) offers a transformative approach to overcome Large Language Model (LLM) limitations and address ESG RegTech challenges. Theoretically, RAG enhances LLMs by grounding outputs in verifiable, current external knowledge from authoritative sources, directly countering hallucination and static knowledge issues (Abdulrahman Bagais et al., 2024; Parikh and Penfield, 2024). This mechanism provides LLMs with context-specific information, improving their understanding of nuanced ESG regulatory questions and complex disclosure requirements (Abdulrahman Bagais et al., 2024; Parikh and Penfield, 2024). Practically, RAG improves accuracy in information extraction by filtering relevant data (Abdulrahman Bagais et al., 2024), supports auditable compliance verification through output traceability (Du et al., 2025), and assists users in navigating voluminous ESG regulatory documents by retrieving targeted information (Parikh and Penfield, 2024). It also offers an agile method for domain-specific adaptation, an alternative to resource-intensive fine-tuning (Page et al., 2021) and enhances automatic question answering from large ESG reports (Sun et al., 2024). Sun et al. (Abdulrahman Bagais et al., 2024) note RAG's utility in the “ESGReveal framework,” while Du et al. (Parikh and Penfield, 2024) identify it for bolstering domain-specific abilities and addressing LLM challenges like hallucination. To further clarify this enhancement, Table 2 systematically outlines the principal limitations of standalone LLMs within ESG RegTech applications and illustrates how the RAG architecture directly addresses these specific challenges.

Table 2. LLM limitations in ESG RegTech and how RAG addresses them

LLM Limitation for ESG RegTech	Evidence from Reviewed Literature	How RAG Addresses the Limitation
Factual Inaccuracies / Hallucinations	(Du et al., 2025; Parikh and Penfield, 2024; Zou et al., 2025)	Grounds generation in verifiable, retrieved external data.
Static / Outdated Knowledge	(Du et al., 2025; Sun et al., 2024)	Enables access to up-to-date information via dynamic external knowledge bases.
Context Window Limits & Processing Costs for Large Docs	(Sun et al., 2024)	Retrieves only relevant segments, making LLM processing more focused and efficient.
Lack of Transparency & Verifiability	(Du et al., 2025; Zou et al., 2025)	Can provide source attribution for retrieved information, enhancing auditability.
Domain Specialization Gap / Nuance Misinterpretation	(Abdulrahman Bagais et al., 2024; Page et al., 2021; Webersinke et al., 2021; Zou et al., 2025)	Provides specific domain context at query time, improving relevance and precision without full model retraining for every update.

4.3. Identified Gaps in RAG for ESG Regulation

This subsection addresses RQ4 by evaluating the current state of peer-reviewed research on the explicit application, empirical evaluation, or robust conceptualization of RAG for ESG RegTech, compliance, and financial regulation, and subsequently identifies key underexplored areas and significant research gaps.

4.3.1. Current State of Research on RAG for ESG RegTech in Reviewed Literature

Current research on Retrieval Augmented Generation (RAG) for ESG RegTech is nascent, despite RAG addressing well-documented challenges. While Sun et al. (Abdulrahman Bagais et al., 2024) directly propose RAG for enhanced ESG information extraction, and Du et al. (Parikh and Penfield, 2024) highlight it for mitigating hallucination and improving domain-specificity in financial NLP, other reviewed papers primarily discuss the problems RAG is designed to solve. These problems include managing voluminous ESG data (Sun et al., 2024), ensuring factual accuracy and currency (Braun and Clarke, 2006; Du et al., 2025), achieving domain adaptation (Page et al., 2021), and building trustworthy AI (Zou et al., 2025). Crucially, none of the nine reviewed papers provide detailed empirical evaluations, comparative benchmarking, or mature case studies of RAG systems specifically deployed for ESG regulatory compliance tasks, such as verifying adherence to CSRD, SFDR, or EU Taxonomy. Current discussions remain largely conceptual or focus broadly on ESG data processing, rather than specific regulatory compliance verification using RAG.

4.3.2. Key Underexplored Areas and Significant Research Gaps

Our analysis of the literature reveals several key research gaps concerning Retrieval Augmented Generation (RAG) in ESG RegTech. A significant gap is the limited empirical validation and benchmarking of RAG systems across diverse ESG regulatory contexts, with existing literature acknowledging RAG's potential but lacking rigorous empirical studies (Abdulrahman Bagais et al., 2024; Parikh and Penfield, 2024). Secondly, there is a scarcity of publicly available, domain-specific datasets and standardized evaluation metrics for RAG in ESG RegTech (Abdulrahman Bagais et al., 2024; Sun et al., 2024; Zou et al., 2025). Thirdly, research on optimal retrieval strategies and knowledge base curation for complex, multi-jurisdictional ESG regulations remains insufficient, with specifics on knowledge base structuring and the potential of knowledge graphs underexplored (Abdulrahman Bagais et al., 2024; Parikh and Penfield, 2024). Fourthly, underdeveloped frameworks exist for ensuring and assessing the explainability, trustworthiness, and auditability of RAG-driven ESG compliance outputs, despite noted LLM opacity concerns (Braun and Clarke, 2006; Du et al., 2025; Zou et al., 2025). Fifth, there is minimal exploration of effective human-RAG collaboration models and interfaces tailored for ESG compliance professionals and regulators (Sun et al., 2024) Sixth, the critical dimensions of data protection and cybersecurity when processing sensitive corporate information are significantly underexplored. Seventh, the practical business case for adoption—including implementation expenses, organizational resource needs, and return on investment—lacks academic investigation, posing a barrier to real-world deployment. Finally, addressing ethical considerations, including bias and fairness, in RAG systems applied to ESG regulatory decision-making requires more focused attention, beyond general LLM ethical concerns (Braun and Clarke, 2006; Zou et al., 2025)

5. Discussion

The qualitative thematic analysis of the peer-reviewed literature from Web of Science and Scopus, as detailed in Section 4, compellingly substantiates the imperative for Retrieval Augmented Generation (RAG) as a critical technological evolution for ESG Regulatory Technology (RegTech). The synthesized evidence reveals a clear consensus regarding the significant challenges that standalone Large Language Models (LLMs) face when applied to the increasing sophistication and dynamism of ESG regulatory demands (RQ1, RQ2). Recurrent themes in the analyzed literature highlight LLM limitations such as their static "knowledge cutoff" (Du et al., 2025), their potential for factual inaccuracies or "hallucinations" (Parikh and Penfield, 2024; Zou et al., 2025), and their difficulties in processing voluminous and nuanced regulatory texts (Abdulrahman Bagais et al., 2024; Sun et al., 2024). These shortcomings create a significant operational and reliability gap, as the ESG regulatory domain demands exceptionally high accuracy, timeliness, and verifiability—characteristics that standalone LLMs inherently struggle to guarantee.

RAG emerges from this analysis as the most promising architectural solution to bridge this identified gap (RQ3). Its core principle of dynamically retrieving relevant, up-to-date information from curated external knowledge bases—such as regulatory documents and corporate disclosures—before LLM generation directly addresses the key LLM limitations (Abdulrahman Bagais et al., 2024; Parikh and Penfield, 2024). The literature conceptualizes RAG as a vital mechanism to enhance factual grounding, provide access to current regulatory intelligence, improve contextual understanding of complex compliance requirements, and bolster transparency through potential source attribution. This positions RAG not merely as an incremental improvement but as a necessary advancement for developing trustworthy and effective AI in ESG RegTech. While the explicit application and empirical validation of RAG specifically for ESG regulatory compliance are still nascent in the reviewed literature (RQ4) (Abdulrahman Bagais et al., 2024), the qualitative evidence strongly supports its unique value in addressing the distinct challenges of this high-stakes domain, moving beyond general ESG insights towards auditable compliance

6. Conclusion

This study asserts the necessity of Retrieval Augmented Generation (RAG) for improving ESG financial regulation and compliance, based on a systematic literature review. Our findings indicate that while Large Language Models (LLMs) can process extensive ESG data (RQ1), their limitations—including factual inaccuracies, static knowledge, and poor contextual understanding—make them unreliable for high-stakes ESG RegTech (RQ2). We synthesize evidence demonstrating that RAG addresses these shortcomings by grounding LLM responses in current, verifiable data, thus enhancing accuracy, timeliness, relevance, and transparency in AI-driven ESG analysis (RQ3). Nevertheless, dedicated peer-reviewed research on RAG's explicit application and empirical validation for ESG regulatory compliance remains nascent (RQ4). Our primary contribution is an evidence-based argument for RAG's critical role in current AI challenges within ESG regulation, positioning it as a pivotal technology for reliable, auditable, and intelligent ESG RegTech solutions. Future research should move towards empirical validation of RAG systems across diverse ESG regulatory contexts, the development of practical deployment strategies, and the establishment of robust frameworks for explainability, trustworthiness, data protection, and ethical application.

REFERENCES

Abdulrahman Bagais, F. M., Dandan, S. M. M., Alsaleh Algazo, F. A., ALDahabi, Z., Alshareef, A., & Kh Barakat, S. A. (2024). Global ESG integration in governance: A systematic review of accountability frameworks, challenges, and strategic innovations. Retrieved May 31, 2025, from https://www.jisem-journal.com/

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative research in psychology, 3(2), 77-101.

Du, K., Zhao, Y., Mao, R., Xing, F., & Cambria, E. (2025). Natural language processing in finance: A survey. Information Fusion, 115, 102755.

Fan, W., Ding, Y., Ning, L., Wang, S., Li, H., Yin, D., ... & Li, Q. (2024, August). A survey on rag meeting llms: Towards retrieval-augmented large language models. In Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining (pp. 6491-6501).

Huang, A. H., Wang, H., & Yang, Y. (2023). FinBERT: A large language model for extracting information from financial text. Contemporary Accounting Research, 40(2), 806-841.

Kim, M., Kim, S., Kim, Y., & Moon, J. (2024, August). Analyzing the financial impact of ESG news sentiment on ESG finance trends. In 2024 International Conference on Platform Technology and Service (PlatCon) (pp. 95-100). IEEE.

Lee, H., Kim, J. H., & Jung, H. S. (2025). From corporate earnings calls to social impact: Exploring ESG signals in S&P 500 ESG index companies through transformer-based models. Journal of Cleaner Production, 501, 145320.

Li, W., Liu, W., Deng, M., Liu, X., & Feng, L. (2025). The impact of large language models on accounting and future application scenarios. Journal of Accounting Literature.

Lim, T. (2024). Environmental, social, and governance (ESG) and artificial intelligence in finance: State-of-the-art and research takeaways. Artificial Intelligence Review, 57(4), 76.

Alahira, J., Mhlongo, N. Z., Falaiye, T., Olubusola, O., Daraojimba, A. I., & Oguejiofor, B. B. (2024). The role of artificial intelligence in enhancing tax compliance and financial regulation. Finance & Accounting Research Journal, 10, 241-251.

Olaiya, O. P., Adesoga, T. O., Pieterson, K., & Qazeem, O. (2024). RegTech Solutions: Enhancing compliance and risk management in the financial industry. GSC Advanced Research and Reviews, 20(2), 8-15.

Ogunyemi, F. (2023). AI-Powered Carbon Accounting: Transforming ESG Reporting Standards for a Sustainable Global Economy. International Journal of Multidisciplinary Research in Science, Engineering and Technology, 6(4).

Parikh, P., & Penfield, J. (2024). Automatic Question Answering From Large ESG Reports. International Journal of Data Warehousing and Mining (IJDWM), 20(1), 1-21.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. bmj, 372.

Sun, Z., Satapathy, R., Guo, D., Li, B., Liu, X., Zhang, Y., ... & Goh, R. S. M. (2024, December). Information extraction: Unstructured to structured for esg reports. In 2024 IEEE International Conference on Data Mining Workshops (ICDMW) (pp. 487-495). IEEE.

Thimm, H., & Rasmussen, K. B. (2024). ChatGPT discovery of green image damaging information for large production companies. Journal of Cleaner Production, 478, 143978.

Vaghefi, S. A., Stammbach, D., Muccione, V., Bingler, J., Ni, J., Kraus, M., ... & Leippold, M. (2023). ChatClimate: Grounding conversational AI in climate science. Communications Earth & Environment, 4(1), 480.

Villacampa-Porta, J., Coronado-Vaca, M., & Garrido-Merchán, E. C. (2025). Impact of EU non-financial reporting regulation on Spanish companies' environmental disclosure: a cutting-edge natural language processing approach. Environmental Sciences Europe, 37(1), 1-33.

Webersinke, N., Kraus, M., Bingler, J. A., & Leippold, M. (2021). Climatebert: A pretrained language model for climate-related text. arXiv preprint arXiv:2110.12010.

Moodaley, W., & Telukdarie, A. (2023). Greenwashing, sustainability reporting, and artificial intelligence: A systematic literature review. Sustainability, 15(2), 1481.

Yang, J. Y., Chi, R. H., Wu, C. C., Chen, L. J., Lin, W. M., Hu, H. W., & Cheng, H. R. (2024, June). EcoSmartGuide: Language Learning Model and Retrieval-Augmented Generation-Based Platform for Streamlined Environmental, Social, and Governance Information Access and Report Generation. In 2024 IEEE 6th Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS) (pp. 343-347). IEEE.

Zou, Y., Shi, M., Chen, Z., Deng, Z., Lei, Z., Zeng, Z., ... & Zhou, W. (2025). ESGReveal: An LLM-based approach for extracting structured data from ESG reports. Journal of Cleaner Production, 489, 144572.